IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF PENNSYLVANIA

| | | |
|---|---|---|
| AMERICAN CIVIL LIBERTIES UNION, *et al.*, | ) ) ) | |
| Plaintiffs. | ) ) ) | |
| v. | ) ) | Civil Action No. 98-CV-5591 |
| ALBERTO R. GONZALES, in his official capacity as Attorney General of the United States. | ) ) ) ) | |
| Defendant. | ) ) | |

**DECLARATION OF PAUL MEWETT**
**IN SUPPORT OF DEFENDANT'S MOTION TO COMPEL**

I, Paul Mewett, hereby declare and state as follows:

1.   I am a Principal at CRA International ("CRAI"), a company where I also consulted for a number of years before taking up a full time post. I qualified as an Electronics Engineer in 1977, and then spent a period as an IT consultant specifically focusing on advising companies on new and evolving technologies including Internet search techniques and data mining. I then established a company in 1998 called IBNet which I subsequently took public in 2000 on the London Stock Exchange. IBNet specialized in using advanced spidering technology to identify brand specific content on the Internet for global businesses. My current duties at CRAI are within the Business Intelligence and Investigations practice and include creating and using Internet intelligence gathering tools to support major fraud investigations. This can involve email tracking, suspect identification, trend analysis, monitoring of evolving issues or advising on Internet strategies. I work closely with, and support, Forensic Accountants and Computer Forensics specialists. I submit this declaration in support of defendant's motion to compel.

2.   At the direction of defendant, Bert Craven, a consultant working for CRAI, and I assisted defendant with the task of obtaining complete, navigable copies of each of twenty-five websites belonging to the plaintiffs in this action. In order to obtain these copies, Mr. Craven and I employed six different "spidering" tools, including one that I have been informed was suggested by plaintiffs: WinHTTrack, WebWhacker, Web Grabber, WebSeeker, wget, and Onfolio.

3.  Spidering tools work by starting with the main, home or index page of a website and following all the links to other files and images, and other associated content. The spidering tool then recreates the directory structure of the site on a local computer and stores all the retrieved files in their correct location. Spidering tools can be effective in obtaining websites that contain only static HTML formatted files. Some websites are a collection of HTML formatted files, linked together using hyperlinks, which are passed to a user's browser as they are requested. The HTML, when displayed, may contain images which are automatically retrieved by the user's browser when it encounters HTML describing an image. It may also contain other embedded media or refer to other types of linked file which are also automatically retrieved in the same way. Spidering tools can make a complete copy of static HTML websites. However, linking techniques such as client-side code execution may hide the destination of a link from a spider, and thus prevent spidering tools from obtaining complete copies of some static HTML websites.

4.  Spidering tools, however, are not an effective method for generating a copy of a dynamic website. A dynamic website is a program that produces output in HTML format in response to requests. The output for an identical request may vary for a virtually unlimited number of reasons, including the time or date of the request, the user's browser, the user's preferred language, the user's location, the user's identity, etc. The output may also contain information from one or more databases, which themselves may be constantly updated by user interaction, such as store inventory, forums, chat rooms, etc. For this reason, spidering tools cannot make reliable or complete copies of a website containing dynamic content.

5.  In light of the shortcomings of spidering tools, neither Mr. Craven nor I was able to be certain that we obtained a complete copy of any plaintiff's website. It does appear, however, that we may have obtained comprehensive copies of three websites: wildcatintl.com, epgn.com, and citylights.com. An examination of the remaining twenty-two sites revealed the following: ten used some techniques that obfuscated links; sixteen included content that appeared to be from one or more databases; and ten included user-authored content, *i.e.*, forums, chat rooms, etc. In our opinion, spidering tools cannot be used to obtain complete copies of the remaining plaintiffs' websites, and it cannot be guaranteed that the initial three copies attempted are complete.

6.  I have been informed that plaintiffs' counsel has asserted that plaintiffs' websites are equally accessible to the defendant as they are to plaintiffs. This is incorrect. As discussed above, almost all plaintiffs' websites contain either hidden links and/or dynamic content. In order to obtain a functional copy of a dynamically-generated page, one would need to obtain not only the static content of any one web page as it exists in a moment of time, but one would also need to obtain the program that generates the page as well as the inputs that program uses, such as data or script files. The programs and data underlying plaintiffs' dynamically-generated web pages are not accessible to the public.
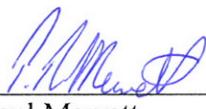
7.  Although spidering tools do not offer an effective method to obtain complete copies of the plaintiffs' websites, in my opinion, plaintiffs themselves are highly likely to be able to make complete, navigable copies. Foremost, it is highly likely that most

plaintiff websites are located on commercial web hosting services, and most commercial web hosting providers make working backup copies at regular intervals whether the web author or content provider is aware of it or not. In addition, most website authors or content providers, or their web designers or programmers, keep backup copies of all program and content materials. In any event, even if plaintiffs do not use commercial web hosting companies that make backup copies, or maintain backup copies of their program and content materials, plaintiffs can make a complete copy of their websites directly from their web servers. It is not difficult to create a backup copy of a website, and, as discussed above, either plaintiffs or their agents almost certainly already have backup copies of their websites in their possession. By their very nature, these backup copies are easily stored, transported, and transferred. Thus, plaintiffs would not face a significant burden if they were to be required to provide the defendant with backup copies of their websites.

8. I have been informed that plaintiffs' counsel has asserted that making a backup copy would require creating a "document." However, making a copy of a website does not require creation of a document. Rather, a backup copy is functionally no different than photocopying an existing paper document. If a particular document was requested from the plaintiff they would not send the original but instead make a photocopy. In the same way, if a copy of a website was requested and one were not readily available, then the required files could merely be copied onto a DVD drive connected to the web server or network, or alternatively a tape backup device could be run to again collate and backup all the required files. This is a standard procedure and not one that requires specialist computer equipment or knowledge.

9. I have also been informed that plaintiffs' counsel has asserted that it would be difficult for plaintiffs to provide full backup copies of their websites, because those copies could only be viewed through the use of the same "server," out of the allegedly more than one hundred servers in existence, that was used in creating the copy. I understand plaintiffs' counsel to use the term "server" to mean the software that is used to interpret, compile, and execute the programs that are used to create and maintain plaintiffs' websites. In our review of the twenty-five plaintiff's websites, however, we have identified only about eight unique software programs that have been used to interpret, compile, and execute the software used to create plaintiffs' websites. In fact, it should be possible for a single physical machine to run each of the twenty-five plaintiffs' websites, using the software applicable to each website. Indeed, unless the plaintiffs have dedicated hosting services, their websites already are co-located with many other websites that use a variety of different software programs to operate those various websites.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Paul Mewett                    22 Nov 05